REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
16-05-2013

2. REPORT TYPE
R&D Status Report (Quarterly)

3. DATES COVERED *(From - To)*
20-02-2013 to 16-05-2013

4. TITLE AND SUBTITLE
Rapid Parallel Screening for Strain Optimization

5a. CONTRACT NUMBER
HR0011-12-C-0062

5b. GRANT NUMBER
NA

5c. PROGRAM ELEMENT NUMBER
NA

6. AUTHOR(S)
Chuck Merryman

Tony Yee

5d. PROJECT NUMBER
NA

5e. TASK NUMBER
NA

5f. WORK UNIT NUMBER
NA

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
J. Craig Venter Institute
9704 Medical Center Drive
Rockville, MD 20850

8. PERFORMING ORGANIZATION REPORT NUMBER
HR0011-12-C-0062.4

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DARPA, MTO
Prog: Living Foundries ATGC
675 N Randolph St
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
DARPA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NA

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES
NA

14. ABSTRACT
Progress has been made on identifying biosensors that will be used to report on the fermentation yields of industrially relevant biological compounds. Screening of the desired chemicals was completed previously. Microbes that can utilize these compounds as their sole carbon source have been isolated and archived. Genomic DNA from 97 isolates was prepared for sequencing last quarter. Sixty-nine samples were submitted to sequencing and sixty-five of those passed quality controls and have been sequenced and assembled. Genome annotation is ongoing. About 10 of the sequenced genomes have been fully annotated and more are completed each day. Bioinformatic analysis of completed genomes has provided candidate transcription factors based on their localization adjacent to or within appropriate catabolic gene clusters. Experimental work has produced and verified two vectors for screening these transcription factors and clusters: 1) a modular assembly vector with a reporter, and, 2) a yeast TAR cloning shuttle vector for transferring catabolic clusters to *E. coli*.

15. SUBJECT TERMS
NA

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Tony Yee
UU	UU	UU	UU	4	**19b. TELEPHONE NUMBER** *(include area code)* 301 795 7133

Rapid parallel screening for strain optimization

Report Title:	Quarterly R&D Status Report
Report Number:	HR0011-12-C-0062.4
Reporting Period:	February 20, 2013 to May 16, 2013
Contract No.:	HR0011-12-C-0062
Performing Organization:	J. Craig Venter Institute 9704 medical Center Drive Rockville, MD 20850 USA
Principal Investigator:	Chuck Merryman

<u>Abstract</u>

Progress has been made on identifying biosensors that will be used to report on the fermentation yields of industrially relevant biological compounds. Screening of the desired chemicals was completed previously. Microbes that can utilize these compounds as their sole carbon source have been isolated and archived. Genomic DNA from 97 isolates was prepared for sequencing last quarter. Sixty-nine samples were submitted to sequencing and sixty-five of those passed quality controls and have been sequenced and assembled. Genome annotation is ongoing. About 10 of the sequenced genomes have been fully annotated and more are completed each day. Bioinformatic analysis of completed genomes has provided candidate transcription factors based on their localization adjacent to or within appropriate catabolic gene clusters. Experimental work has produced and verified two vectors for screening these transcription factors and clusters: 1) a modular assembly vector with a reporter, and, 2) a yeast TAR cloning shuttle vector for transferring catabolic clusters to *E. coli*.

Sponsored by
Defense Advanced Research Projects Agency
Microsystems Technology Office (MTO)
Program: Living Foundries: Advanced Tools and Capabilities for Generalizable Platforms (ATCG)
Issued by DARPA/CMO under Contract No: HR0011-12-C-0062

Table of Contents:

Section

Summary

In total, 108 compounds have been used for enrichment culture and 85 compounds produced colonies (when used as the sole source of carbon and energy). In this reporting period, we have completed sequencing candidate organisms and are nearly finished with genome assembly and bioinformatic analysis of these microbes. In agreement with expectations and earlier phylotyping results, pseudomonas species predominate. Close scrutiny of a fully assembled and annotated genome has revealed the presence of more than 100 transcription factors and at least 20 biodegradation clusters that also contain transcription factors. Six of these transcription factors are prime candidates for recognizing the corresponding carbon source (chrysin) because the assosicated degradation pathways are rich in enzymatic activities for processing the chemicals ring structures. Constructs for cloning and evaluating transcription factors and gene clusters have been made and verified.

Introduction

The overall goal in this contract is to link cell-based production to cell survival and thereby make the engineering of new microbial strains that produce industrially relevant biochemicals routine. Recent synthetic biology techniques can make billions of variant cells. Although, many potentially informative mutants are easily made, product yield can only be determined in a few of these. The majority of industrially relevant biomolecules are not chromophores, naturally discernible, or otherwise easily detected. Nevertheless, genetic circuits are capable of linking chemical production to discernible signals such as growth or color intensity. Such a system would allow numerous mutants and mutant combinations to be examined quickly. Genetic circuits needed to screen mutant populations in parallel rely upon the availability of an appropriate biosensor that activates a reporter gene in a product dependent fashion. In this project, genes for two-component and one-component signaling systems (that respond to industrially relevant biomolecules) are identified using microbial growth assays, sequencing, and quantitative PCR (qPCR). To demonstrate that such sensors can be used to maximize product yield, one sensing system will be further engineered. We will reformat this sensor so that it drives expression of a reporter such as an antibiotic resistance marker. This sensor/resistance cassette, and a biosynthetic pathway capable of producing the molecule to which the sensor responds, will be placed within a heterologous host that does not have an overlapping pathway. Basal synthesis of the targeted chemical (by the orthogonal biosynthetic pathway) activates the sensor and increases transcription of the resistance marker (i.e. reporter). In other words, the fermentation product is also the sensor ligand and thus, biosynthesis drives production of the reporter and cell survival. Antibiotic levels in the media will be adjusted so that basal product yield, and hence basal marker activity, is insufficient for survival. Targeted, genome-wide and barcoded alterations to the host genome will then be installed. Variants with better and better chemical production survive by virtue of their ability to withstand increasing antibiotic challenge.

Methods, Assumptions and Procedures

Over the next two quarters we were slated to experimentally validate 5-10 candidate transcription factors by measuring their response to exogenously supplied substrates (*i.e.* task 5). Some of the work necessary for this task was moved forward due to the delay in sequencing. Thus we have assembled and verified two vectors for screening transcription factors and associated catabolic clusters. The modular assembly vector places a tetA reporter under control of the T5 promoter and the transcription factor under constitutive expression. The T5 promoter is fused with putative operators that respond to the transcription factor. The initial construct responds to cumate. We have assembled the vector so that other transcription factors and operators can be swapped into the construct to replace elements in the test vector that are specific to cumate. All that is necessary is amplification and assembly of 3 DNA fragments. Likewise, we have created and used a TAR cloning vector for transferring catabolic clusters to *E. coli*. A pseudomonas nicotinic acid catabolic cluster and transcription factor were captured to test this vector. For this approach, candidate transcription factors are identified by growth of *E. coli* on the target substrate (i.e. by complementation).

The original plan was to complete sequencing last quarter and finish analysis this quarter. Sequencing (and thus analysis) was delayed because the manufacturer had backordered library construction kits. Sequencing has now been finished and the corresponding genomes are being analyzed. During this delay, we realized that our original plans for bioinformatic analysis could be improved by identifying transcription factors that were adjacent to or within catabolic gene clusters. Largely, this approach is analogous to antiSMASH, which focuses on identifying anabolic pathways. Thus, in our new approach, we prioritize transcription factors based on their clustering with catabolic enzymes. For each chemical structure, enzymes likely needed for its degradation can also be predicted to further identify prime candidates. This additional analysis has now been implemented and will inform the design of differential qPCR experiments and downstream verification.

Results and Discussion

One-hundred-eight compounds have been used as the sole carbon and energy source for microbial growth (M9 salts supplemented with a target chemical). Colonies were produced for eighty-five compounds (~80%).Sixty-nine clones were selected. Sixty-five of these passed quality controls, have been sequenced, and the respective genomes assembled. Bioinformatic analysis has been improved and is partially completed. Annotation, transcription factor identification, and catabolic gene cluster identification have been finished for ~15% of the sequenced and assembled clones. An assembled, annotated, and auto-analyzed genome was scrutinized by hand and revealed 20 biodegradation clusters with corresponding transcription factors. Six of these are prime candidates for recognizing the corresponding carbon source (chrysin) because the assosicated degradation pathways are rich in enzymatic activities for processing the chemicals ring structures. This indicates that we will be able to experimentally focus on transcription factors with a high probability of responding to a specified target chemical. We have assembled and verified two vectors for screening transcription factors and associated catabolic clusters. These will be used to verify that a specified transcription factor responds to a target chemical.

Conclusions

The results indicate that a chemical made by one organism is likely to be used as food by some other microbe. Bacteria typically utilize the most efficient carbon source available (glucose often being the preferred substrate). More exotic carbon sources are generally subject to catabolite repression and systems for their utilization are activated after preferred carbon sources are exhausted. Besides catabolite repression, sensors are often employed so that the appropriate degradation pathway for a non-preferred carbon source is activated. Our sequencing results have identified organisms rich in transcription-factor based sensors that are integrated with appropriate catabolic gene clusters. The next step is to experimentally test individual sensors for their ability to respond to a specified carbon source. This will set the stage for downstream work on the overproduction of such molecules. Identification and experimental validation of specific sensors is indispensable but current results are promising: 1) most target chemicals (~80%) readily produce microbial growth, 2) colony morphology, etc., suggests that different target chemicals resulted in the isolation of different microbial species, 3) automatable procedures can be used if high-throughput screens are needed in the future, 4) about 70% of recovered organisms are from the genus *Pseudomonas*, consistent with the commonplace identification of sensors and degradation pathways within this genus when anthropogenic chemicals are used as targets, 5) fully sequenced and annotated genomes are rich in transcription factors, and, 6) in the first fully analyzed case, several transcription factors are part of catabolic gene clusters that might degrade the corresponding chemical.

Statement of Work Task List:

- Task 1 (Phase I, Year 1, Months 0-3): Completed (please refer to report HR0011-12-C-2.1)
- Task 2 (Phase I, Year 1, Months 4-9): Completed. Sixty-five isolates have been sequenced.
- Task 3 (Phase I, Year 1, Months 10-12): Almost completed. The genomes sequenced above have been assembled and ~15% of these finished annotation and improved bioinformatics analysis.
- Task 5 (Phase II, Year 2, Months 13-18): Initiated. Construction and testing of two-vectors needed to experimentally validate transcription factors was moved forward due to sequencing delays.

Planned Activities for the Next Reporting Period

During the next reporting period we will finalize bioinformatics analysis select candidate genes, and verify them by qPCR and their response to target chemicals in a heterologous host.

Program Financial Status

	Planned Expend	Actual Expend (Cumulative to Date)	% Budget Completion	At Completion	Latest Revised Estimate	Remarks
Task 1	$59,251	$59,251	100%	$59,251	$59,251	Completed
Task 2	$69,229	$69,229	100%	$69,229	$69,229	Completed
Task 3	$124,706	$89,496	72%	N/A	$124,706	N/A
Task 4	$255,817	$0	0	N/A	$255,817	N/A
Task 5	$255,817	$30,334	12%	N/A	$255,817	N/A
Cumulative	$764,820	$248,310	32%	N/A	$764,820	N/A

There is no management reserve or unallocated resources.

Based on the currently authorized work:

- Is current funding sufficient for the current fiscal year? Yes
- What is the next fiscal year funding requirement at current anticipated levels? The budgeted amount for Year 2 of the project is $396,905.25.
- Have you included in the report narrative any explanation of the above data and are they cross-referenced? Not applicable; current funding is sufficient for the current fiscal year.